

PARA	Unitary	1
imilarity matrix	missmatch penalty	1.00
	gap penalty	0.33
	gap size penalty	0.33
	outoff score	2.00
	alpha	2.00

	SEARCH STATISTICS			Total Elapsed
CORES:	Mean	Median	Standard Deviation	Time
randomization group	487	0	0.00	00:00:00

number of residues : 2785
number of sequences searched : 1
number of matches : 00:00:00

The scores below are sorted by initial score.
Significance is calculated based on initial score.
100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init.	Opt.	Sig.	Frame
1. m33987	TOIG OF: m33987	check: 7224	2785	487	558	0.00	0
1. US-10-069-434-4 (1-911)	TOIG OF: m33987	check: 7224	from: 1 to: 2785				
m33987	TOIG OF: m33987	check: 7224	from: 1 to: 2785				
Locus HUMCAIX	2785 bp mRNA linear	PRI 31-OCT-1994					
DEFINITION Human carbonic anhydrase I (CAI) mRNA, complete cds.							
ACCESSION M33987							
VERSION M33987.1							
KEYWORDS carbonic anhydrase I.							
ORGANISM Homo sapiens (human)							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.							
REFERENCE 1 (bases 1 to 2785)							
AUTHORS Lowe, N.; Brady, H.J.; Barlow, J.H.; Sowden, J.C.; Edwards, M. and Butterworth, P.H.							
TITLE Structure and methylation patterns of the gene encoding human carbonic anhydrase I							
JOURNAL Gene 93 (2), 277-283 (1990)							
MEDLINE 9103039							
PUTMED 2121614							
COMMENT Original source text: Human EBV transformed SH B cell line DNA, and cDNA to mRNA.							
Since no intron sequences were provided this entry is treated as if originating from an mRNA.							
Draft entry and computer-readable sequence for [Unpublished (1990)] kindly submitted by N.Lowe, 09-MAY-1990.							
Author address: N.Lowe Department of Biochemistry University College London Gower Street, London WC1E 6BT, U.K. E-mail: UCBMARM@BUCLID.UCL.AC.UK@CDNYVM.CUNY.EDU.							
FEATURES	Location/Qualifiers						
source	1..2785						
	/organism="Homo sapiens"						
	/mol_type="mRNA"						
	/db_xref="taxon:9606"						
gene	/map="8q13-q22"						
	1..2785						
	/gene="CA1"						
TATA_signal	874..879						
mRNA	/gene="CA1"						
	902..2165						
	/gene="CA1"						
	/note="carbonic anhydrase I mRNA (alt.)"						
mRNA	902..1939						
	/gene="CA1"						
	/note="carbonic anhydrase I mRNA (alt.)"						
CDS	1048..1833						
	/gene="CA1"						
	/note="carbonic anhydrase I (EC 4.2.1.1)"						
	/codon_start=1						
	/protein_id="AAA51910-1"						
	/db_xref="GI:179793"						
	/db_xref="GDB:G00-119-047"						
	/translation="MA3PDWGYDDKLYPANQNNNOSPVDIKTSCTKHDTSLKPISVSYNAPATAKEIINVGHSFHVNPFSDSYRFLQHFHEWPKLQKHEGSEHTYDGKYSAEHHAHNSAKTSSLAQDLAGTVGLMKRGAQNPKLQKVLDALQAIKTKGRAPTFNDESTLIPSSLDWTYPOSILTHPLYESVTWICKESISVSSEQLAQFRSLSVNGDNAVPMQHNRPQPLKRTVRAFS"						
ORIGIN	2785	September 22	2004	12:44	2000	N	Choi, 7002
	m33987						

Initial Score	Optimized Score	Significance	Matches	Mismatches	Gaps
= 487	= 558	= 0.00	= 299	= 0	
Residue Identity = 62%	Conservative Substitutions = 599				
Residue Identity = 61%	Conservative Substitutions = 558				
TCTCTATTAGAGCTGTTTCCACAGAGTAGTGTAAAGAAACTGGATTTCAGAAGGAA	X CGGAATTC---GGCTGAG--				
960 970	980 990	1000 X 1010	1020		
- - - T T C C A C C C G - A G - G G C C A T G T C A G G G T C A G - C T G G G G A T A C C G C A G C A P A C G S T C C T A T T C A C T	10 110	70 80			
A A A G G R A A C T C A G T G A A G A G T A G T G C C A G T G A C A A T G G C C T G A C A A T	1030 1040	1050 1060	1070 1080		
90 100	110 120	130 140	150		
G G A A G G A A - T T T T C C T T A T T G C T G A T G C G A A T C T C C A A T T G A G A T T A A A C C A A A G A G T G A A G	1110 1120	1130 1140	1150 1160		
A T A T G A C M C T C C C T C G A C C A C T T G A T G A C A G A G A C C A A A C C A A T T G A G A T T A A A C C A A A A C C A A	1170 1180	1190 1200	1210 1220		
A C A T G A C A C C T C T G A A C C T T G A T G A C A G A A T T G A G A T T A A A C C A A A C C A A	1240 1250	1260 1270	1280 1290		
160 170	180 190	200 210	220		
C C A T T C T C T C A A T T T G A C T T G A T G A C A G A G A C C A A A C C A A T T G A G A T T A A A C C A A A G C G G	1290 1300	1310 1320			
G C A T T C T C C A T T A A T T T G A G A C A G A A C C A A T T G A G A T T A A A C C A A A C C A A	1330 1340	1350 1360			
300 310	320 330	340 350	360		
C T A C A G G T P A C G G A G G T T C A C C T T C A C T G G G G T C C G G T G A T G A C A C C G G T C C G A G A C A T A G T G A G	1370 1380	1390 1400	1410 1420		
C T A C A G G T P A C G G A G G T T C A C C T T C A C T G G G G T C C G G T G A T G A C A C C G G T C C G A G A C A T A G T G A G	1430 1440	1450 1460	1470 1480		
370 380	390 400	410 420	430		
A G T G A G C T A T G C T G G A G S T C C A C T T G G T T G C T A C T G G A A T T G A C A A A C T G G A T T G A G G O A G	1490 1500	1510 1520	1530 1540		
A G T C A A T T T C G C C A G G T T C A G G T T G C G A T T T C A G T G A T T G G T T G A G G T T G G C C C A A - C T G C A	1550 1560	1570 1580	1590 1600		
440 450	460 470	480 490	500		
C T C A T G A A C C A G A T G G A C T T G G A C A T T A A A G G T A A A C A C T G G A C A T A T T T G A C T T A	1610 1620	1630 1640	1650 1660		
G A A - A G T A T C T G A C T G C C T C C A A G G T T A A A C C A A G G C C C A A G G G C C C A A T T T G A C T T C T G A	1670 1680	1690 1700	1710 1720		
510 520	530 540	550 560	570		
A A A G T T A C - T G A C A C T T G G A C T T G G A C A T T A A A G G T A A A C A C T G G A C A T A T T T G A C T T A	1730 1740	1750 1760	1770 1780		
T G T C T G C T C C T C C T C T A T C C T G G A C T T G G A C A T T A A A C C T A A C T G G T T C T G A C T T C C T C T G A	1790 1800	1810 1820	1830 1840		
T G T G T A C T T G G T T A C T C T G G A C T T G G A C A T T A A A C C T A A C T G G T T C T G A C T T C C T C T G A	1850 1860	1870 1880	1890 1900		
730 740	750 760	770 780	790		
T G T G C A C A G G G G G G T C A A G G A G G T T T C T G T A G A G G T C A C - T C T G T C - A C C C A - - G G C T G G A	1910 1920	1930 1940	1950 1960		

```
> O <
| O | O IntelliGenetics
O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4
```

Results file 069434_4_x_m33987_inv.res made by spaula on Wed 22 Sep 104 12:49:06-EST.
Query sequence being compared: US-10-069-434-4' (1-911)
Number of sequences searched: 1
Number of scores above cutoff: 1

Results of the initial comparison of US-10-069-434-4' (1-911) with:
File : m33987.seq

100-

-

U 50-

M -

B -

E -

R -

O -

P 10-

C -

S -

E 5-

Q -

U -

E -

N -

C -

S -

PARAMETERS

Similarity matrix Unitary
Mismatch penalty 1
Gap penalty 1.00
Gap size penalty 0.33
Cutoff score 0
Randomization group 0

SEARCH STATISTICS

Scores: Mean 68 Median 0 Standard Deviation 0.00 Total Elapsed 00:00:00.00

Times:

CPU

Length:

m33987

Number of residues:

Number of sequences searched:

Number of scores above cutoff:

1

The scores below are sorted by initial score.
Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt. Length	Score	Sig. Frame
1. m33987	TOIG of: m33987 check: 7224	2785	68	410 0.00 0
1. US-10-069-434-4' (1-911)	TOIG of: m33987 check: 7224 from: 1 to: 2785	2785	68	410 0.00 0
TOIG of: m33987 check: 7224 from: 1 to: 2785				
LOCUS HUMCAIX	Human carbonic anhydrase I (CA1) mRNA linear	PRI 31-OCT-1994		
DEFINITION Human carbonic anhydrase I (CA1) mRNA, complete cds.				
ACCESSION M33987				
VERSION M33987.1				
KEYWORDS carbonic anhydrase I.				
ORGANISM Homo sapiens (human)				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE 1 (bases 1 to 2785)				
AUTHORS Lowe,N., Brady,H.J., Barlow,J.H., Sowden,J.C., Edwards,M. and Butterworth,P.H.				
TITLE Structure and methylation patterns of the gene encoding human carbonic anhydrase I				
JOURNAL Gene 93 (2), 277-283 (1990)				
MEDLINE 91033039				
PUBMED 2121614				
COMMENT Original source text: Human EBV transformed SH B cell line DNA, and cDNA to mRNA.				
E-MAIL Since no intron sequences were provided this entry is treated as if originating from an mRNA.				
FEATURES Draft entry and computer-readable sequence for [Unpublished (1990)] kindly submitted by N.Lowe, 09-MAY-1990.				
Source Author address: N.Lowe Department of Biochemistry University College London Gower Street, London WC1E 6BT, U.K. E-mail:UCBCNMAR@EUCLID.UCL.AC.UK@CUNYVM.CUNY.EDU.				
1..2785				
Location/Qualifiers /organism="Homo sapiens"				
/mol type="mRNA"				
/db_xref="taxon:9606"				
/map="8q13-q22"				
1..2785				
/gene="CA1"				
874..879				
/gene="CA1"				
902..902				
/gene="CA1"				
2165..2165				
/gene="CA1"				
902..902				
/note="carboxylic anhydride I mRNA (alt.)"				
mRNA				
902..902				
/gene="CA1"				
1048..1048				
/gene="CA1"				
1048..1048				
/note="carboxylic anhydride I (EC 4.2.1.1)"				
/codon_start=1				
/protein_id="AAAS1910.1"				
/db_xref="GI:179793"				
/db_xref="GDB:G00-119-047"				
TRANSLATION="MA\$PDGYDDKNGPBDWSKLYPIANGNNQSPVDIKTSETRHTDSLKPISVSYPNATAKEIINVGSPFNEIDNDRNLVKGPFSDSYRLFDFHEHGSTMTHGSSEHTVQCKVYSAEHLVHMSAKYSSLAEEASKDGLAVIGVLMRGEANPKLQKVLDLQAUKKGKAPTFNDPSTLPSLDFWTVPPGSUTHPPLYESTWILKESISVSSEQIQLAQPSLILSRVNEGNAPVQHNNRPQTPLGRVTVRA\$"				
ORIGIN				
M33987 Length: 2785	September 22, 2004 12:44	Type: N	Check: 7224	..

Initial Score = 68 Optimized Score = 410 Significance = 0.00
 Residue Identity = 50% Matches = 535 N mismatches = 329
 Gaps = 201 Conservative Substitutions = 0

```

          X      10
          TTTTC-CCCTGGCATT-GGG
          ||||| | | | | | | | | |
          CAGTTTCCCTCTGAGCTGATTAATCCACACCCCCAACCAACACTCCATGGTTCTCACATCTGG
          670   690   700   710   X   720   730
          20      30      40      50      60      70      80
          GAAAGCTATG--CCAGTGGTCAAGACTCTGGAGGTGAGGGGATCHA--CCTGAGTCCAGAGT
          ||||| | | | | | | | | | | | | | | | | | | | |
          GGCCTACT-ATGTACCCACT-CTAAATCACACACGGCGCACATCACATGGAGGCCCTCCC
          740   750   760   770   780   790   800
          90      100     110     120     130     140     150
          GGAGGCTGC--AATTAGCCAATTGTAACCCTGCCCCTCCAGCTGGTGGTGGAGACTT-CA
          ||||| | | | | | | | | | | | | | | | | | | | |
          AAAGCCGCCAAATTATGCAA-ATTAT-TCAAATGTAAC---CCACCCCTTGAGACG
          810   820   830   840   850   860   870
          160     170     180     190     200
          CAGAAAGCTGCTTCACCCCTCC-GC-TGTG---CA---CAGGAG--AC---TGC---GAATA
          ||||| | | | | | | | | | | | | | | | | | | | |
          TACATAGTGCCTCATCCCTCCAGCTGTGTGTTCAAGGTTCACTGTTGCTCT
          880   890   900   910   920   930   940
          210     220     230     240     250
          G-GCCAGC-----TGTGAGAGCTGATGT-TTATAGTTGCTTAAACA---TCCATGTGACA
          ||||| | | | | | | | | | | | | | | | | | | | |
          GTGGGA-GCCCTCTCATTCATCAGCTGTTCCACAGGTTAG-TGAAAGACGTTCA
          950   960   970   980   990   1000   1010
          260     270     280     290     300     310     320
          CTCT-CAAGAGAGCTGGAACTGTGAAAG-AAGAACAGGAT-ATGTCAGTAGTCCAGGATGGT-GAAGCAG
          ||||| | | | | | | | | | | | | | | | | | | | |
          CTRGCAAG-AGAAAAGAAAACCTAAGAACTAAGAA---GATAATGGAAAGT---GATAATGGGATATGATG
          950   960   970   980   990   1000   1010
          330     340     350     360     370     380     390
          AGACATAGTCAAAATTGTGAATCAGTTGTTACCTTGTGTTAAATGGAA---TCCAAGATGTGAG-T
          ||||| | | | | | | | | | | | | | | | | | | | |
          ACAAAATGTCCTAACATGGAAAG-CTGTAT-CCATTCG-CATGGAAATACCAATCCCTGTT
          1080  1090  1100  1110  1120  1130  1140
          400     410     420     430     440     450     510
          AATCTT--TTGCACT-TGGAATTAGGTACCAATCTGAA-AAACA---CTCC-CAAGACAGC-C
          ||||| | | | | | | | | | | | | | | | | | | | |
          GATCTAAACAGTGGAAACCAACATG-ACACCTCTGAAACCTATGTTGAAATTGAGGAAACAGATAACCGAT
          1150  1160  1170  1180  1190  1200  1210
          460     470     480     490     500     560     570
          AGTCTTCTGT--TCA-TAGCTGGCT---CA-ACAAGCTGGTATTGTCGA---ATTCCAGTG
          ||||| | | | | | | | | | | | | | | | | | | | |
          AGCCAAAGAAATTATGTTCTGCAATTGTTCCATTGTAATTTGAGGAAACAGATAACCGAT
          1220  1230  1240  1250  1260  1270  1280
          520     530     540     550     560     620     630
          -AACACATGG---AGCTCTG-CAGC-ATA-GCTACTCATACTATGTGCTGGAGCCGTTGTCATCA
          ||||| | | | | | | | | | | | | | | | | | | | |
          CTGAAAGGGTGTCCATTCTGAGAGCTACAGGCTTCACTGTTTCACTTGTGTTCA
          1290  1300  1310  1320  1330  1340  1350
          580     590     600     610
          GC--GG-----ACCCCGT-GAAGGTG-AA---CCMCCGCPAACTT-GTAGCTTCAGTGAGGA
          ||||| | | | | | | | | | | | | | | | | | | | |
          GCATGTTCAAGAACATACAGTGGATGGATGAAATTCTGCGC-AGCTTCACGTGTTCA
          1360  1370  1380  1390  1400  1410  1420
          640     650     660     670     680     690
          COACCCGCGAAGCTGAACTGATTGTTCTGTCATCAAAG-TCAACATT-GAAGGAATGGCGTGTGCT
          ||||| | | | | | | | | | | | | | | | | | | |
        
```